Research Paper 27 | **2014**

A DRAGON EATING ITS OWN TAIL: PUBLIC INFORMATION ABOUT POLLUTION IN CHINA

Chiara RAVETTI, Yana JIN, Mu QUAN,
Zhang SHIQIU, Tim SWANSON

A DRAGON EATING ITS OWN TAIL: PUBLIC INFORMATION ABOUT POLLUTION IN CHINA *

Chiara Ravetti § Yana Jin‡ Mu Quan‡ Zhang Shiqiu‡ Tim Swanson ¶

This draft: 9 September 2015

Abstract

This paper examines how a government that controls both public information and pollution emissions can exercise discretion in its choice of pollution signals, and thus influence pollution responses in the population. We develop a model of the government's optimal decision about pollution emissions and information about them, and we apply its results in the context of the information distortions and the adaptation choices of households in Beijing, China. We proceed in three stages. First, we use a simple signal extraction model to motivate why a government may choose to distort information about pollution. Then, we perform a time series analysis on air pollution announcements, to test empirically if the information signal in Beijing is biased when compared to an alternative measure from the US Embassy. This analysis indicates that public information is systematically modified, as predicted by our model. Finally, using data from an original household survey, we examine the effect of the distorted public signal on agents' behavior, and find that those who rely on public media controlled by the government are significantly less responsive to pollution peaks, demonstrating the impact of governmental control over information.

Keywords: Air Pollution; Signal Extraction; Information; Averting Behaviour.

JEL Classification: H41, Q53, Q56, Q58, I13.

---

*The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under the grant agreement n^o 266992 (GLOBAL IQ). We benefited from the comments of participants to the International Economics Association World Congress, of the Swiss Program of Environmental and Energy Economics, of the Italian Society for Climate Sciences and of the Italian Association of Resource Economics and of the Young Swiss Economist Meeting. We thank Richard Baldwin, Nicolas Berman, Anthony Venables, Ugo Panizza, Michael Stimmelmayr, Arun Jacob and Rahul Mehrotra, for their insightful comments and suggestions. The usual disclaimer applies.

§Corresponding author. Oxford Centre for the Analysis of Resource Rich Economies (Oxcarre), Department of Economics, Manor Road, Oxford, OX1 3UQ Email: *chiara.ravetti@economics.ox.ac.uk*

‡Peking University, College of Environmental Sciences and Engineering, Beijing, China.

¶Department of International Economics and CIES, Graduate Institute of Geneva (IHEID), Maison de la Paix, Ch. Eugène-Rigot 2,CH-1211, Geneva, Switzerland.

1. Introduction

Public sector information is a valuable public good that can greatly influence the expectations of economic agents. Economists have long recognized that information can be incomplete and costly to acquire (Stiglitz, 2000) and that governments can reduce asymmetries through public information on issues ranging from meteorological forecasts, to inflation targets, to official statistics (Morris and Shin, 2002). Moreover, public information plays a central role in ensuring institutional quality (Brunetti and Weder, 2003; Besley and Prat, 2006; Islam, 2006). However, public sector information can be double-hedged sword: a government can have incentives to provide limited or distorted information and to redirect public opinion, especially when it can largely control the media (Williams, 2009).

This paper investigates the role of public information in the context of air pollution in China. This country is an emblematic example of a developing nation where information is strictly controlled by the government, and that suffers from severe urban air pollution.[1] We focus on the health risks posed by air pollution because, as for many other environmental hazards, information about this problem is difficult to acquire individually: gathering data is costly and requires specialized technology (monitors) and knowledge (epidemiological research). Thus, pollution information is often publicly provided by the state. Here, however, we question if a government would report pollution information accurately, whenever it can exercise large control over public information and the media.

This research belongs to the rich literature on the importance of public information for economic outcomes. Much research has focused on the role of public information on macroeconomic outcomes (e.g. Mèon and Minne, 2014 on exchange rate regimes, Baeriswyl and Cornand, 2010 on central banks signals) and on government policies (Besley and Burgess, 2002; Gavazza and Lizzeri, 2009), with fewer works on environmental effects (Kennedy et al., 1994). Indeed some papers find that governments might have an incentive to misreport data strategically (Michalski and Stoltz, 2013). There are a number of papers examining specifically air pollution information in China (see Andrews, 2008, Yuyu et al., 2012 and Ghanem and Zhang, 2014, Stoerk, 2015) that find evidence of data manipulation with respect to air pollution. Differently from the existing literature, this paper provides a theoretical motivation about the driving forces behind information distortions, and exploits unique data from the capital of China, Beijing, to uncover the specific mechanism behind this publicly-enforced imperfect information signal. We first analyse why a government might choose optimally to misreport information about air pollution, and then we look at the evidence about this misreporting and at its impact on household behaviour.

First, we develop a model of state control over public information, looking at the trade-off faced by a

[1] Air quality is a well-known problem for Chinese cities: 7 out of 10 of the most polluted cities in the world are in China, Beijing being one of them (Asian Development Bank, 2007). The cost of ozone and particulate matter concentration in China was estimated for 2005 around 112 billion USD (Matus et al., 2012).

government between higher output, and thus higher pollution, against public health costs and labour force participation. In the short run, information influences the ability of households to respond to pollution peaks, while in the long run it affects their choice to work in polluted cities. For the government, the benefit of signalling low pollution is that it can maintain a cheap and complacent labour force in urban areas, at the cost of higher health expenditures. This simple model predicts that, for a government, it is optimal to introduce a downward bias in the pollution signal (declaring that pollution is lower than in reality) for increasing levels of pollution. This effect can become even stronger if the informational environment changes. As a corollary, households who rely on this distorted signal, and are not able to update their expectations correcting for the bias, systematically adopt less self-protective measures against pollution peaks.

Then we test empirically these hypotheses in Beijing. The capital of China offers a unique setting to analyse information signals about air pollution, because there exists an alternative source of data about air quality besides the official Chinese one: a US Embassy measure. Confronting the two information sources, we find that the Chinese government appears to create a systematic downward bias that increases with pollution, especially around emission levels where it can more strongly affect public opinion. A 1% increase in the pollution measured by the US monitor is associated with a 0.8% increment in the downward bias in the pollution signal, and this more than triples around key thresholds.

Next, we examine the implications of these distorted pollution signals at the household level. The hypothesis here is that those agents who rely on public information sources (publicly owned media like TV, radio and newspaper) are directly influenced by public signals and thus less capable of responding to pollution peaks. Numerous papers have explored the socio-economic mechanisms of self-protection against environmental risks.[2] There are a number of studies specifically on the role of information about environmental hazards in developing countries, where the flow of information is often less efficient (Madajewicz et al., 2007, Jalan and Somanathan, 2008, Somanathan, 2010). Overall, these studies find that information can positively affect the response to environmental risks. Instead in this paper, using an original household survey from different districts in Beijing, we find that households who rely on the government signal are significantly less likely to adopt more self-protective measures during pollution peaks.

This paper broadly contributes to a fundamental economic question, namely ‘why are developing countries bad at providing public goods?’ (Besley and Ghatak, 2006). We provide evidence for one possible explanation of this puzzle, by looking at this channel of public information control. Our contribution to the debate is twofold: firstly, we find robust empirical evidence on the signal’s bias in China, supporting the mechanism of information manipulation described theoretically. A government has less of an incentive to provide the environmental public goods (such as clean air)

[2]Economic agents can allocate resources or efforts to risk-reduction (Ehrlich and Becker, 1972), especially when insurance markets are incomplete and in the presence of intangible values, such as good health (Simmons and Kruse, 2000, Cook and Graham, 1977). Empirical research finds that people combine market and non-market strategies to avoid uncertain damages (Simmons and Kruse, 2000, Talberth et al., 2006, Whitehead, 2005).

since it can control through the media the public perception of this issue. Secondly, this research considers different adaptation choices and averting behaviours of households in response to the government announcements. This provides original household-level evidence about the deleterious impact of distorted public information.

The paper proceeds as follows: section 2 presents a simple conceptual framework about government and households' interactions through information; section 3 analyses empirically the data about Beijing's air pollution index; section 4 presents the household survey and the empirical findings about averting behavior with different sources of information. Section 5 concludes.

2. Model

We build a simple conceptual framework to explain why a government may optimally choose to bias information signals. We use a signal-extraction model, with the signal being provided by a self-interested government that can control the media and thus bias pollution announcements. Households rely on these public announcements to form their expectations about air quality. The government can decide to create a wedge between expectations and true pollution, in order to attract a complacent workforce to the city and produce more output. However, by doing so, it reduces the capacity of households to self-protect against health risks, so the bias is costly in terms of public health expenditures. This framework shows that it might be an equilibrium solution for the government to declare pollution to be *lower* than its actual levels, if a fraction of people cannot update their beliefs and correct for the bias. As a consequence, urban dwellers that fully rely on the government signal become less responsive to pollution peaks.

2.1. The government's problem

The government of a large productive city maximizes its own political support function, V, which depends on the profits made by firms, which can lobby for loose environmental regulations, and on the dissatisfaction of citizens with health damages caused by air pollution. Therefore, the government needs to support productive activities, which make use of pollution and labour as inputs, but also to maintain an acceptable level of public health, to avoid discontent about the high cost of airborne diseases. The political support function is

$$V = \Pi - cDN \tag{1}$$

where Π is the aggregate profit of firms, N the number of workers and D health damages, and c a parameter capturing the weight assigned to public health by the government. The trade-off for the government is between favouring the production of output to generate profits and the actual state of public health. These are defined, respectively, as

$$
\begin{aligned}
\Pi &\equiv \pi(p,N) & \partial\Pi/\partial p \geq 0 \qquad \partial^2\Pi/\partial p^2 \leq 0 \\
N &\equiv n(w) - kE(p) & \partial N/\partial E(p) < 0 \\
D &\equiv dp - aE(p) & \partial D/\partial E(p) < 0
\end{aligned}
\tag{2}
$$

where p is pollution, w the urban wage, and $E(p)$ perceived pollution. The interpretation of $n(\cdot)$, k, d and a is detailed below.

In order to produce any output, the city needs workers, N, who must be attracted to a polluted urban area, where they can provide their labour. Overall, the number of workers willing to be actively employed in a polluted city depends on the perception of pollution, $E(p)$, and on the wage, through a function $n(w)$.[3] In this model, working conditions in the city matter. In particular, a perception of excessive air pollution reduces by a factor k the number of people willing to work in the city for the current wage, or reduce the number of workers able to provide their labour under existing conditions (due to inclement working conditions, for example). [4]

If the government can exert some control over public information about pollution, it will have an incentive to announce that air quality is better than it truly is, in order to keep a large and functioning workforce active in the city. At the same time, though, the government loses public support as health damages for urban dwellers rise. These damages increase by a factor d with true pollution, p.[5] However individual agents can reduce pollution damages using some precautionary measures, whose effectiveness depends upon perceived pollution, $E(p)$, and translate into lower health damages with an effectiveness captured by the parameter a. The crux of this model is that economic outcomes do not depend only on true pollution, but also on the *perception* that people have of pollution, $E(p)$. The government can make announcements about air quality in order to influence people's expectations, and this has real consequences for the economy.

[3] Wages as exogenous in this setting, and are determined by market forces and not by the government.

[4] For further references about location and working choices under different life quality attributes, see the hedonic valuation literature, as in Clark et al. (2003), Chay and Greenstone (2005) or Bayer et al. (2009).

[5] Here the parameter d captures the morbidity effect of an increase in a given pollutant, and the costs that the government must bear on account of such (via social insurance for example). In the epidemiological literature, a positive relationship between air pollutants and adverse health outcomes has been identified consistently across a variety of studies in developed countries (Ren and Tong, 2008). In the developing world research is scarcer, but to date a number of studies elicit exposure-response functions specific to developing countries. See, for example, El-Fadel and Massoud (2000) for Lebanon, Sakulniyomporn et al. (2011) for Thailand, Braga et al. (2001) for Brazil, Foster and Kumar (2011) for India. In mainland China, the key epidemiological studies are summarised in a comprehensive meta-analysis by Aunan and Pan (2004), with few more recent cohort studies, such as Qiu et al. (2012) or Zhang et al. (2012).

2.2. *Pollution and information*

How do households extract information about the actual level of pollution, given a government's announcement? There are two distinct groups of agents within the subject population, those who are relatively well-informed and those who are uninformed. The difference between the two groups is one of degree: the informed group have relatively better access to sources of information other than the government (own observation, production information, other outside sources), while the uninformed have only the government reports on pollution levels as a source of information.

We firs discuss the information-processing mechanism that applies to the relatively informed part of the population. Agents alone cannot fully evaluate air quality, but they have some prior beliefs about its characteristics. Also, they receive the government signal, which could be biased, and update their expectations based on this new information.

The prior about pollution in this economy is

$$p = p^n + \theta \quad with \quad \theta \sim N(0, \sigma_\theta^2) \tag{3}$$

where p^n is the "natural" level of pollution, given by the geographic conformation and location of the city, and θ captures are emissions shocks deriving from production. True pollution levels coming from emissions are unknown to individual agents, but the distribution of pollution shocks is generally known. In effect, agents have some knowledge concerning the prevailing levels of pollution in a given locality, but do not have exact knowledge of the emissions occurring on a given day. For simplicity, we set p^n equal to zero without loss of generality, and thus we can use p and θ interchangeably to indicate pollution.

The government can determine how much extra emissions to allow or not into the economy, θ. Then it releases an announcement A about the level of emissions, which can also include a bias β. This is

$$A = p + \beta \quad with \quad \beta \sim N(\beta_0, \sigma_\beta^2) \tag{4}$$

Again, relatively informed households do not know the actual bias of the government β, but have expectations about its mean value and variance. The uninformed households in the population simply take the announcement by the government as fact. [6]

2.3. *Aggregating and updating expectations*

The informed agents form their updated expectations by solving a signal extraction problem: knowing the distribution of pollution shocks and that of the government bias, using current and

[6] This set-up is analogous to an output-gap model of an economy with a central bank deciding on inflation targeting and announcements. Similarly, here the government has control over real variables - pollution shocks - and nominal ones - announcements. For an example of a similar model in the context of a central bank and inflation, see Moscarini (2007).

past observations, they can update their expectations about i) the government bias and ii) current pollution levels. The expected bias is then a precision-weighted sum of the previous signals that people have received, given the distribution of pollution and announcements

$$E(\beta) = \left(\frac{\beta_0}{\sigma_\beta^2} + \frac{T\overline{A}}{\sigma_\theta^2}\right) / \left(\frac{1}{\sigma_\beta^2} + \frac{T}{\sigma_\theta^2}\right) \tag{5}$$

where $\overline{A} = \sum_{t=1}^{T} A/t$ is a weighted average of all past announcements up to time T, the present moment. Clearly the higher the variance of announcements, the lower the weight that people assign them relative to their prior about the bias in forming their expectations.

Expected pollution emissions, then, are obtained by updating the prior about pollution shocks with the government announcement, adjusted for the expected bias.

$$E(\theta) = \left(\frac{A - E(\beta)}{W^2}\right) / \left(\frac{1}{\sigma_\theta^2} + \frac{1}{W^2}\right) \tag{6}$$

where W^2 is the variance of $A - E(\beta)$, the bias-adjusted announcement.[7] If agents can update their beliefs for a sufficiently long period of time, their expectations concerning announcements will converge to the true government bias.

Now we turn to the impact of announcements on the uninformed part of the population. In a country where the media is largely controlled by the government, a fraction λ of the population receives information exclusively from the government and thus is incapable of updating expectations concerning the bias. When aggregating the beliefs over the whole population, on average only $(1-\lambda)$ of the population is able to update their expectations concerning the government's bias, leaving the uninformed part of the population to accept the government's announcement as actual pollution. This means that, on average, only $(1-\lambda)$ of the bias is corrected out of the announcement, yielding in aggregate the following expression

$$E(\theta) = \left(\frac{A - (1-\lambda)E(\beta)}{W^2}\right) / \left(\frac{1}{\sigma_\theta^2} + \frac{1}{W^2}\right) \tag{7}$$

This is the overall population's expected level of pollution, when receiving announcements from a government that controls the flow of information to a substantial portion λ of that population. When the government controls most of information outlets, λ is close to 1 and most agents would consider only the announcement to inform their expectations. Therefore, in the presence of information controls, expectations would never converge to the true value of pollution.

Simplifying the notation, we can define the aggregate expected pollution as

$$E(\theta) = \frac{\theta + \lambda E(\beta)}{z} \tag{8}$$

[7] Similarly to DellaVigna and Kaplan (2007), this is $W^2 = \left(1/\sigma_\beta^2 + (T-1)/\sigma_\theta^2\right) / \left(1/\sigma_\beta^2 + T/\sigma_\theta^2\right)^2$.

where $E(\beta) \to \beta$ over time and $z = W^2 \left(\frac{1}{\sigma_\theta^2} + \frac{1}{W^2} \right)$ is the variance deriving from pollution, bias and announcements.

From this signal extraction problem, we can derive two statements about the role of government biased announcements in pollution expectations.

Proposition 1 - *A government making an announcement A regarding the pollution level θ that includes a bias β affects expectations about air pollution through two channels: directly, through the announcement A, and indirectly, entering the average of past announcements, $\overline{A}$. The first effect exceeds the second, so $\partial E(\theta)/\partial\beta$ >0, more strongly the higher the fraction of the population that is unable to learn about the bias, i.e. $\partial E(\theta)^2/\partial\beta\partial\lambda > 0$.*

Proof - It follows straightforwardly from differentiating equation (7).

This proposition already illustrates the importance of a high degree of control over information, as captured by λ. The greater the control over information flows to the population (high lambda), the more effective the announcements are in pushing people's expectations away from true pollution levels, even in the short run. Over time, through the learning process, the informed part of the population is able to factor out most of the bias, but as long as there is a fraction of people who fully rely on the government announcement, the bias is an effective policy tool to influence pollution expectations. It follows logically that, in this set-up:

Proposition 2 - *Given some fraction of the population that is uninformed about true pollution levels, λ, the government's adoption of a negative bias in its announcements regarding emissions results in expectations about pollution that are lower on average than actual pollution levels, $E(\theta) < \theta$, and viceversa.*

Proof - By inspection of equation 7, and given Proposition 1.

This proposition complements and completes the previous one: whenever the government introduces a negative bias - it announces that air pollution levels are lower than they are in reality - it will produce expectations about pollution that are lower than true pollution levels. How much lower depends on the extent of its control over the flow of information to the population, λ, as stated in Proposition 1.

In turn, this distorted perception about the air pollution problem makes urban workers more complacent about bad air quality, but it also reduces their ability to respond to pollution with self-protective measures.

2.4. Optimal bias and pollution

Next, we examine the optimal policy choices by the government, given its control over information in the city. The government maximizes its value funciton (1) subject to (3). The choice variables are emissions θ and the bias β embedded in the announcements about pollution A.

Ultimately, the government faces a trade-off related to air pollution, as this is useful for production, but creates health damages. At the same time it faces another trade-off regarding the introduction of a bias in pollution information, because while this keeps the population happy, it limits people's capacity to avert pollution hazards. The equilibrium choice of pollution and bias depends on the optimal balancing of these different forces. The government solves the problem backwards, considering expectations about pollution through the labour supply and health function, maximizing its support as shown in equation (1).

The optimal bias and optimal pollution emissions in equilibrium are generally

$$\theta^* = \frac{a}{kd}\, n(w) + \frac{1}{cd}\, \pi_N - \frac{2a}{ckd^2}\, \pi_\theta \tag{9}$$

$$\beta^* = \left[\frac{dz-a}{kd}\, n(w) - \frac{1}{cd}\, \pi_N - \frac{dz-2a}{ckd^2}\, \pi_\theta \right] \frac{1}{\lambda} \tag{10}$$

where π_N and π_θ are the derivatives of profits with respect to the number of workers and pollution, respectively (proof in the Appendix).

These two optimal choices for the government are inter-related, but with some significant differences. The most important distinguishing feature is that the optimal bias β^* depends on λ: the less control over information that the government can exercise (i.e. the lower the λ), the greater the bias it needs to introduce to achieve the same impact on expectations.

Otherwise, the optimal bias is closely related to the level of optimal pollution, as we can see when we re-state the expression for optimal bias as a function of optimal pollution:

$$\beta^* = \frac{1}{\lambda} \left[\frac{(cdn(w) - \pi_\theta)z}{ckd} - \theta^* \right] \tag{11}$$

It is seen that the relationship between the optimal bias and optimal emissions is composed of two effects: directly, through θ^* (the last term of the expression); and also indirectly through the impact of emissions on the profit function, π. This direct effect registers the extent to which the government has control over the information environment, and the incentives to make use of this control. The indirect effect registers the extent to which the government values pollution-induced profits (π_θ) over pollution-induced health damages *cdn* in the economy.

This latter effect demonstrates the basic reason that a government in control of both pollution and information has incentives to distort the latter. Households may respond to pollution by withdrawing involvement in labour markets where production occurs, resulting in loss of profits. A government with control over information regarding pollution has the ability to neutralise such responses, at the cost of increased pollution-sourced health damages. The government will have an incentive to do so to the extent to which it values production-sourced profits over production-sourced health damages, electing to suffer the cost consequences of a less well-informed and so less-responsive population.

2.5. *Optimal bias - response to changes in pollution and information*

We now wish to see how the optimal informational bias moves in response to changes in the optimal level of pollution.

Proposition 3 – *An increase in pollution emissions θ^* affects the bias through two channels: directly, it reduces the bias by a factor $1/\lambda$, and indirectly, by a factor of $\pi_\theta^2/\lambda cd$.*

Proof. It follows from simply differentiating eq. (11) with respect to θ^* and examining the two components of the derivative. □

$$\frac{\partial \beta^*}{\partial \theta} = -\frac{1}{\lambda} - \frac{z}{c\,kd}\underbrace{\left(\frac{\partial \pi_\theta}{\partial \theta}\right)}_{<0} \tag{12}$$

The first term implies that, *ceteris paribus*, increasing pollution exacerbates a downward bias, making announcements of air quality even less alarming. Interestingly this term of the bias function is relatively stable, i.e. it does not change with any parameters of the problem other than pollution levels. The second term is the indirect effect of pollution on information bias, and its sign is ambiguous. If pollution has diminishing marginal returns to the profit function, the second derivative of π is negative and can counterbalance the direct effect. In addition, the second term is also dependent on other parameters of the model, most interestingly the amount of variation in the information environment. So, the overall impact of optimal pollution on the optimal bias will depend not only on the existing level of pollution but also on the informational environment within which it occurs.

In what follows, we will be interested in investigating how changes in the characteristics of the informational environment can impact the negativity of the bias. The variable that best captures the nature of the informational environment is z, which describes the general amount of variability which the environment contains.

The manner in which the informational context (i.e. z) contributes to the way in which optimal bias is ascertained is described as follows:

Proposition 4 – *At points where the governments' signal is less noisy, namely when z is greater, the optimal bias can be even more negative (provided that $\pi_\theta/cd > n(w)$). This effect is mitigated by the interaction of z with increasing pollution, in which case the interaction effect is positive, i.e.:*

$$\frac{\partial \beta^*}{\partial z} < 0 \qquad \frac{\partial^2 \beta^*}{\partial \theta \partial z} > 0$$

Proof. Differentiating eq.(11) and (12) with respect to the parameter z.

□

The informational environment has an impact on optimal bias in announcements. The general variability in information z influences the optimal choice of the government, and so sudden changes in the informational environment may also shift the bias discontinuously. We will examine this phenomenon in the context of sudden changes in air quality, represented by the crossing of thresholds for air quality indexes.

2.6. *Summary: Optimal announcements concerning pollution*

We summarise the way in which a government might optimally exercise control over both information and pollution, by discussing the form which the optimal announcements regarding pollution levels will take. Overall, the optimal announcement for a government is just the sum of the optimal bias and the optimal emission shock

$$\begin{aligned} A^* &\equiv \theta^* + \beta^* \\ &= \left[\frac{(cdn(w) - \pi_\theta)z}{ckd} + \theta^*(\lambda - 1)\right]\frac{1}{\lambda} \end{aligned} \tag{13}$$

This makes clear how the government's informational policies interact with its environmental ones. First, the impact of informational control on announcements is clear. The lower the level of control the government over the population's beliefs (low λ), the larger the magnitude of the announcement needed to achieve the desired effect. Also, for small values of λ (always smaller than 1), the true pollution shocks are always down-played in the final announcement, as captured by the second term in parenthesis.

Secondly, the way in which relative weighting of objectives figures into policy is captured in the first term of Equation (13). This shows again how the government's optimal announcement of information depends upon a balance of health costs versus profits afforded by pollution; the government issues a distorting announcement away from the true pollution level, θ^* when it places greater weight on profits relative to health damages.

Finally, we have seen that the optimal bias will change in response to changes in both the environment (i.e. optimal pollution) and also with respect to the informational environment (i.e. the variability in information).

Overall, the ability to control part of the information environment does not mean that the government will distort information without constraint. The government must consider how its ability to increase the labour supply by distorting information also neutralises the population's responsiveness to pollution more generally. Removing the working population's incentives to respond to pollution via relocation improves production and profitability, but at the cost of generally removing the population's incentives to reduce the costs of pollution. In the following empirical analysis, we examine how this distortion results in additional costliness in the Beijing economy.

The rest of the paper will test the validity of these results, in the context of public information signals about pollution and the responsiveness of households in Beijing.

3. Empirical analysis

Following the discussion in the previous section, we are now able to formulate three testable hypotheses regarding information distortions within an economy where the government controls both information and air quality.

Hypothesis 1. The government of a country that can control the release of information about air quality is likely to introduce a *negative* bias in pollution warnings that increases as pollution gets higher.

This follows from the analysis of the equilibrium bias in eq. (11) and from the inverse relationship between the optimal bias and pollution highlighted in Proposition 3 in the previous section.

Hypothesis 2. At points where the government is facing changes in the informational environment, it will introduce a larger (negative) bias, but decreasingly so as pollution rises.

This follows from Proposition 4 in the theoretical model.

The first and second hypothesis can be tested in an empirical analysis of the bias that takes the following reduced form

$$\beta = \alpha_0 + \alpha_1 \theta + \alpha_2 z + \alpha_3 \theta z \tag{14}$$

where α_0 is a constant corresponding to $zn(w)/\lambda k$ in eq. (11), α_1 is the coefficient of the direct effect of emissions θ on the bias, namely $1/\lambda$ in the same equation, $\alpha_2 z$ captures discontinuous jumps in the z function, and the interaction term between θ and z captures with the coefficient of π_θ, namely $z/\lambda cdk$ in the model.

According to the predictions of the model, we expect $\alpha_0 > 0$, $\alpha_1 < 0$, $\alpha_2 < 0$, and $\alpha_3 > 0$. In the following paragraph we detail how these elements are defined empirically in a complete empirical model.

Hypothesis 3. Agents who fully rely on public sources of information (the fraction λ of the population) will adopt reduced self-protective measures during pollution peaks, in the case of a downwards bias in the informational environment.

This hypothesis tests one of the key premises of the model, rather than its conclusions: if the role of pollution perception $E(\theta)$ is the one described in section 2.2, the higher the expected pollution, the larger the self-protective behaviour that people should adopt. If the government is indeed reporting lower pollution than in reality, these people will perceive less risk and consequently adopt fewer self-protective measures, inducing higher health damages given the distorted informational environment.

We test the first two hypotheses in this section, with a time series analysis of the government signal about pollution. Then the last hypothesis is tested in the context of a household survey in Section 4.

3.1. Air quality signals

The previous theoretical framework establishes how it can be in the interest of the government to distort information signals. We now turn to the empirical evidence in the case of China. First of all, we examine public air pollution information provided by the government in Beijing, released by the Ministry of Environmental Protection, and we compare it to another source of information, the index provided by the US Embassy. Then, in the next section, we will use a household survey to link this phenomenon to individual behaviour in response to air pollution.

In most countries, public agencies monitor and diffuse information about a city's pollution parameters through Air Quality Indexes (AQI). This service provides a public good that otherwise would be under-supplied at a private level. The format of such indexes in the USA, Canada and the European Union reflects international standards defined by the World Health Organization (WHO, 2005) and conveys information about pollution risk through a simple rating of air quality. The signal needs to be clear and understandable to the general public, thus it does not report the concentration of individual pollutants, but rather a color-coded scale going from green (lowest pollution) to dark red (highest pollution). The index score is then associated with a description of the potential health damages associated with it, to highlight the gravity of the environmental situation.

China has adopted an Air Pollution Index (API) that is largely comparable to the one used by the US Environmental Protection Agency (EPA). Table 1 compares the two indexes, showing that the boundaries for each level of the scale are equivalent. Both the Chinese Air Pollution Index or the US Air Quality index are constructed using the same non-linear algorithm using pollution concentration (eq. 15 below): on a given day, they take the highest index value given by any of the pollutants that compose it.

$$API = max(I_1, I_2, ...I_n) \quad where \quad I_i = (C - C_{low})\frac{I_{high} - I_{low}}{C_{high} - C_{low}} + I_{low} \tag{15}$$

In this formula, C is the concentration of pollutant i, I is the index value, and *high* and *low* indicate the boundaries of each category mentioned in Table1 (Beijing Municipal Environment Monitoring Centre and US-EPA 2006). [8] In terms of health consequences, the two indexes convey the same information: an index below 100 implies very little risk of health damages (and is colour coded with green); as the index rises between 100 and 200 (yellow) and 200-300 (orange), more people can be affected by the pollution; and a signal above 300 is defined as a health alert (dark red), with all the population risking severe health consequences. The two indexes, thus, are meant to convey the same signal and information content to the population, and thanks to the simple color-coding of these broad categories, they can determine the general perception of air quality.

3.2. Air pollution information

We collect four and a half years of daily pollution data about the two air pollution indexes, from August 2008 to January 2013. The Chinese data was retrieved from the Ministry of Environmental Protection, while the US Embassy publishes its data in the form of an hourly message on Twitter. The two raw time series are shown in Fig. 3 and 4 in the Appendix. The mean value over this time period for the Chinese measure is 83 (a score of *Good* according to the Chinese definition, such that air quality is acceptable), while for the US index the average is 160, classified as *Unhealthy* by US standards. The US index reaches peaks as high as 800 points and repeatedly reaches 500 during our time period, while the Chinese index goes above 300 points only in five occasions.

Since the US Embassy data is reported hourly, we take the 24-hours average of this data in order to perform a straightforward comparison between the indexes. Aggregating over time is of course not neutral, and therefore in the robustness section we compare with other possible values. Fig. 1 shows a graphical comparison of the two indexes during a short time period within our dataset.

The two indexes take different values in the same day, and this could be partly due to measurement errors or noise in both signals. However there also seems to be a recurring pattern. The Chinese

[8] While the formula to compute the index is identical between the two countries, the boundaries of concentration are different for NOx and other pollutants, so in the construction of the dataset we consider only those days when particulate matter of 10 micrometers of diameter, PM_{10} is the main pollutant.

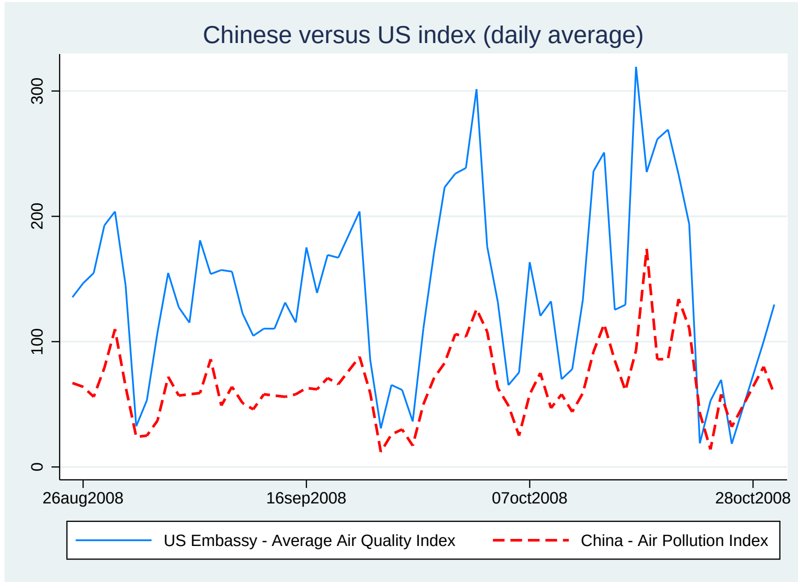

FIGURE 1: MISMATCH BETWEEN CHINESE AND US POLLUTION INDEX

index is not only lower than the US one but is prominently lower in highly polluted days, while during less polluted days the two indexes are more similar. A constant gap between the two time series could be due to different sensitivity of monitors, or even to their orientation towards or away from a polluted street. Therefore we want to explore if there are any systematic patterns that do not just produce a stable difference over time between the two signals, but that could suggest that under specific circumstances the two indexes diverge. We turn in the next section to this analysis.

3.3. Empirical model

In order to gather more formal insights into any systematic manipulation of Beijing's air quality index, we conduct the following time-series analysis. We model the discrepancy between the two indexes as an auto-regressive, moving average (ARMA) process. The two series are stationary according to a simple Dickey-Fuller test, therefore we can apply standard time-series techniques to examine how the two indexes relate.

The dependent variable of the ARMA process is the gap between the two indexes, measured in natural logarithms. The average value over the time period 2008-2013 of the dependent variable is -0.6, indicating that on average the Chinese signal is lower than the US one (see Table 5.). This could include any simple difference between the two indexes due to monitor sensitivity or location of the monitors, but also any government bias. In order to disentangle these elements, we exploit the structure of information signals: while the monitors measure a continuous variable, namely pollutants' concentration, the number that comes out of it has very different meanings to the population, depending whether the number is lower than 100 (green colour code, the air is fine), above 100 (yellow-orange code, air starts being polluted) or above 300 (dark red code, the air is highly hazardous). These steps, as explained before, are merely information devices to inform the population more clearly, and have nothing to do with the original concentration in pollution levels measured by monitors. Therefore, if we observe jumps in the gap between the two indexes around these thresholds, they cannot be due to technical differences, but rather would suggest a government manipulation in order to affect people's perception.

Therefore, we regress the gap between the two signals over the following explanatory variables: i) the US index alone as a proxy for true air pollution - this under the assumption that the US embassy does not have itself any incentive to distort systematically its signal (especially upwards, which would easily cause diplomatic tensions); ii) controls for the 100 and 300 thresholds in the index, as they mark the boundaries between the broad categories of good air quality, and heavy pollution (as shown in Table 1).[9]; and iii) moving average and autoregressive components, to capture persistence and inertia in the pollution processes.

The hypothesis is that, if the government is concerned about people's perception of pollution, it must ensure that the index does not cross the information thresholds too often. It is more effective for the state to change the signal from 101 to 99, since this will imply a color code of green instead of yellow and a signal of good air, than to move the index from 99 to 97, as these are within the same category. This links back to Hypothesis 1 and 2 from the theoretical model: we expect the bias not only to increase as pollution gets higher, but also to display some non-linearity around these thresholds, which represent important changes in the characteristics of the informational environment.

The final empirical specification is the following:

$$G_t = \alpha + \beta I_t^{US} + \sum_{i=1}^{p} \gamma_i T_i + \sum_{i=1}^{p} \delta_i T_i I_t^{US} + \sum_{i=1}^{q} \phi_i G_{t-j} + \sum_{l=1}^{r} \theta_l \varepsilon_{t-l} + \eta_m + \sigma_y + \varepsilon_t$$

where the dependent variable G_t is the gap between the information signal provided by the Chinese government and the US one, namely $G_t \equiv I_t^C / I_t^{US}$, capturing also the bias in the Chinese signal. Each index, I_t^K, is the natural logarithm of the air pollution signal with $K \in \{China, US\}$. T_i is a dummy variable equal to zero below the threshold, and to 1 above it. It captures the importance of informing people with different "regions" of pollution damage: for example an index just above 300 announces to the population that the air is "Heavy Polluted" (dark red), but a small change around this point that brings the index below 300 can strongly affect people's perception. We further interact the thresholds with the US pollution index, to see how the effect diminishes as we get away from the threshold. In order to capture persistence in shocks and in the stock of pollution, we include an autoregressive term, with lags of the dependent variable, and some lags of the error term. The exact number of lags can be determined comparing different models, which is done in the results section.

[9] We also include the threshold 200, just in case there is some discontinuity between yellow and orange coded signals, even if we suspect this will not have a strong effect.

Finally we include month and year fixed effects, to capture anomalous periods in the dataset, such as the Beijing Olympics.

If there was no systematic distortion in the pollution signal, the two indexes should be close and move in tandem, hence the dependent variable will take values on average close to zero and the effect of an increase in the US index should be insignificant. The more discrepancy between the two, however, and the more of a (negative) significant impact we should observe.

3.4. Results

The results of different autoregressive-moving average specifications are presented in Table 3. The negative and significant coefficient of the US index indicates that, whenever pollution (as registered by the US measurement) rises, the effect on the bias is unambiguously downwards: a 1% increase in pollution makes the bias more negative of 0.8%. This result could be foreseen given the graphs showed before: the Chinese pollution index does not increase as much as the US one, so whenever pollution raises, the ratio between the two indexes goes down.

Perhaps more interestingly, beyond this linear trend in the bias, we find that thresholds play an important role in moving the asymmetric relation between the Chinese and the US Index. This gives an indication of potential data manipulation. Both the 100 and 300 thresholds have significant negative coefficients. The effect of pollution raising above 100 or 300 diminishes as pollution increases, as shown by the positive and significant interaction term between thresholds and the US index: in fact, it is easier to downplay pollution levels just above the threshold (e.g. 103 or 302), shifting to the category below, than values largely above the threshold.

The different columns of Table 3 show different specifications of the model in terms of the number of autoregressive and moving average terms.[10] These models can be compared on the basis of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are used for model selection by indicating the relative goodness of fit.[11] Choosing the two models with lowest information criteria, namely the autoregressive models with one or two lags (AR1 and AR2), we compute in-sample forecasts to see which of the chosen models performs best in terms of predictive power. Comparing the mean squared errors of our forecasts (or alternatively the absolute value of the predicted errors), we select the most suitable lag structure among all the ones examined, which is the autoregressive one with one period lag, AR1.

The forecasts for this model are plotted in Figure 5 in the appendix, which shows how the model specified predicts the Chinese bias inside of the sample using previous information and the lags

[10]Here we report only the models that have all ARMA variables significant. more specifications with different lag structures are available upon request.

[11]These criteria are only interpretable as relative measures to compare different models, they have no absolute meaning with respect to goodness of fit, as the classical R squared would have.

structure selected. The model follows closely the fluctuations in the bias, but it still predicts less of a negative bias than the actual observed one.

3.5. Robustness

One important assumption in the previous analysis was that the Chinese index is constructed as an average of various observations over the day. This might not necessarily be the case. Since the US air pollution index is constructed from hourly observations, we can check if the official Chinese air quality index is closer to the US Embassy one when using daily minima and if the bias is still present.[12] The discrepancy between the two, however, persists. A graphical comparison between the US minimum daily value and the usual Beijing index is shown in Figure 2.

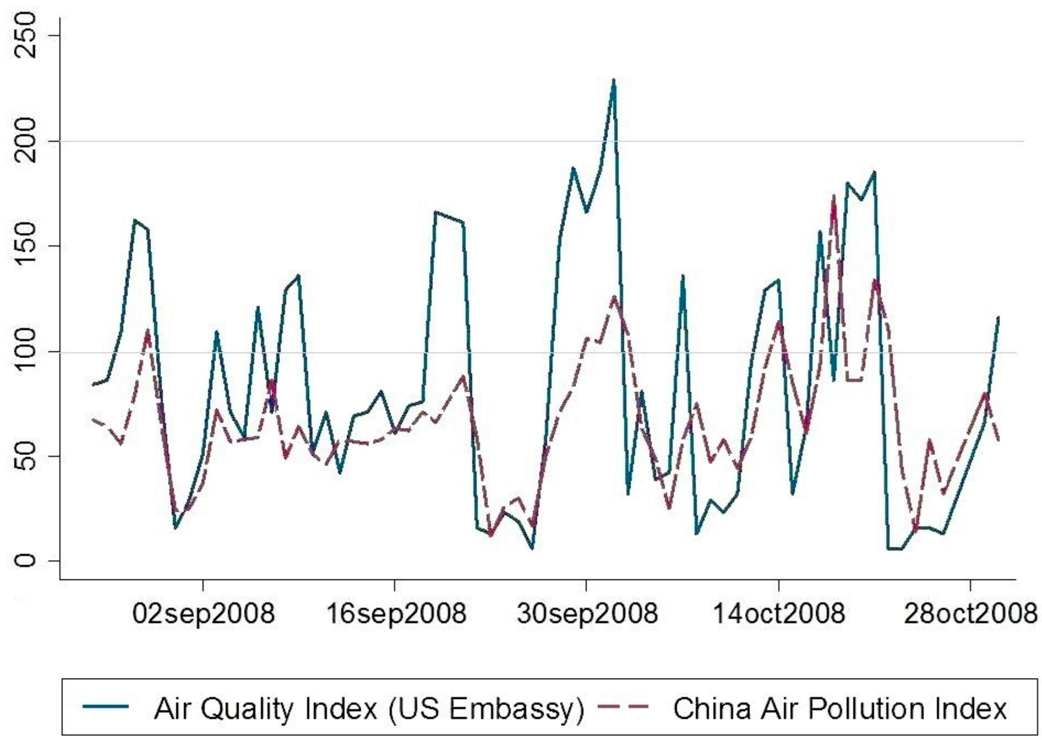

FIGURE 2: INDEX MISMATCH WITH DAILY MINIMA

The results of the ARMA model are robust to different specifications of the bias: when we use the distance of the Chinese index from the minimum of the US index (the most conservative measure we can take), the same results hold both for the linear trend and for thresholds, only with slight changes in the magnitudes of the coefficients (Table 4).

[12]Naturally, the gap between the two indexes is even more pronounced when considering the maximum value in a day of the US measurement (see Figure 6 in the Appendix).

Overall, this analysis robustly shows that the Chinese signal about air quality contains a systematical downward bias, especially around significant thresholds, and this suggests that, as hypothesized in the theoretical model, the government is choosing to misguide popular expectations about pollution. The benefits for the government of declaring that the air is cleaner outweigh the costs. The next section is then dedicated to the analysis of household responses to this distorted air pollution signal, to examine the capacity of individuals to update their expectations beyond this government bias.

4. Households

Individuals who live in Beijing can choose how to respond to the environmental and health hazard presented by air pollution, by virtue of incurring a cost (monetary or in terms of time) for protecting themselves and their families from the damages of pollution. Furthermore, agents receive information signals about air pollution, either from government controlled sources (TV, newspapers, radio), or using their own perception of pollution - for instance observing visibility - or from alternative sources of information, which are still relatively less popular in China, such as the internet. For instance, the US Embassy measurements are freely available via Twitter every hour, and they can even be downloaded on a mobile device as an application. However the internet is restricted in China, and typically only young people are able to access this sort of alternative information using virtual private networks.

In order to examine the behaviour of households with respect to air pollution and information, we collected household data through a survey in urban Beijing, to elicits the expenditure and time allocation to self-protective activities. A set of questions was dedicated to sources of information, in order to identify which groups are more capable of accessing different types of signals during peak pollution days. This allows us to test directly the theoretical model's underlying assumption that some people are more informed and some other instead rely only on the government's announcements. Here we can see if indeed the difference in information sources and in the ability to correct for potential distortions can affect the self-protective measures adopted by the population. For a full description of the survey data and a sample questionnaire, see Ravetti et al. (2014).

4.1. Data

The survey was administered in three districts of Beijing, Haidian, Chaoyan and Dongcheng, for a total of 1672 individuals in 578 households. Due to time and financial constraints, only few districts and neighbourhoods were chosen. The sample selection was designed to represent accurately the total population: we applied probability proportional to size (PPS) at the district and street level and random selection at the community and household level, so that all households in Beijing had equal chances of selection. The sampling probability for a given household was

$$p_0 \frac{[N_H]_{D1}}{[N_h]_{TOT}} * p_1 \frac{[[N_H]_{S1}]_{D1}}{[N_h]_{D1}} * p_2 \frac{1}{[[N_{C1}]_{S1}]_{D1}} * \frac{x}{[[[N_H]_{C1}]_{S1}]_{D1}} = c$$

where the first term captures the probability of the district being chosen, the second that of a street being chosen (both done with PPS); the third and fourth term represent the probability of a community and a household in a given street being chosen, both through random selection. Overall, the sampling design yielded a constant probability for a household in any district, street or community to be selected. Again we refer to Ravetti et al. (2014) for the details of the sampling methodology.

The questionnaire inquired in detail about i) the socio-economic characteristics of the household; ii) various habits and self-protective behaviours: wearing masks, reducing time outdoor, changing means of transportation, doing preventive health checks, using air purifiers; iii) how the family gathered information about air pollution; and iv) health of family members and particularly airborne diseases, cost of illness and insurance.[13] The respondents (one per household) could only answer for themselves and for close family members who spent most of the time in the household. The average household size in the sample is around 3, which is reasonable given the one child policy. Comparing various demographic characteristics of the sample with the Statistics Bureau of Beijing, the survey is in line with the characteristics of the total population, so the sample can be considered representative.

The data from the household survey vary over three dimensions: across individuals, within households and somewhat over time. For the time variation, which will be important in defining our dependent variable of averting behaviour, respondents had to recall their averting behaviour choices in periods of extreme pollution peaks and over the rest of the year, which provides variation between normal times and extreme pollution events. To distinguish between extreme and normal times, the respondents needed to recall the two worst episodes of air pollution in Beijing in the previous year, and to locate them in time. In the year before the survey, in fact, there were two major pollution alerts during hazardous pollution days.[14] Thus we define as extremely polluted days these extreme situations. Self-protective behaviour should increase following signals of extreme circumstances, even if normally a person is not too careful about air pollution. We test this in the following sections.

[13] The survey was administered in Chinese.

[14] Only 66 % of respondents had noticed the extremely polluted haze days in Beijing, indicating that even in those cases there was no widespread information about the pollution risks.

The use of recall data to introduce this time dimension is not free from limitations, but it gives a sense of people's variation in behaviour vis-a-vis pollution peaks.

4.2. Stylized Facts

Cost of illness. First of all, this dataset is useful to examine the characteristics of people in Beijing facing the air pollution hazard. The first thing to notice is that the private cost of illness is quite high: only from airborne diseases, the average annual expenditure including medical costs, medicines and foregone wage is more than 3000 yuan, almost a month of average salary (see Table 6 in the Appendix). [15] This indicates that air pollution imposes some significant costs on households, and thus there can be scope for rational self-protective behaviour, to try to reduce these health expenses.

Self-protective behaviours. The survey captures in detail time-use, weekly exposure to outdoor pollution and self-protective behaviours. Table 7 in the Appendix illustrates the characteristics of the averting choices considered. There are five possible actions undertaken to reduce health risks. *Reducing time outdoor* captures the decision to spend less time outside for leisure and for exercise purposes, so it defines changes in the free-time schedule of a person. This behaviour is much more frequent during extremely polluted days (77 % of the sample declared they adopted it). *Transport change* instead implies moving from means of transportation with high exposure to pollution (such as. walking or biking) to relatively safer ones, such as using a car or a taxi. The percentage of people adopting this behaviour in extremely polluted days doubles, but this is not a common strategy that people choose or can afford. *Masks* are also a relatively infrequent behaviour, adopted by less than 20 % of the sample. The questionnaire differentiates between common paper masks and more sophisticated ones, as there exist more expensive masks on the market that can filter more efficiently particulate matter, but only few people used the higher quality masks (2-3 % of the sample). Finally, the questionnaire includes two different, more expensive long term strategies: *preventive medical checks* and *air purifiers*. The former considers those check-ups of the respiratory system for which the person had to pay some medical costs. The price paid for preventive health checks annually ranged from 10 to 15,000 yuan. Buying an air purifier is an extremely infrequent behaviour, as air purifiers can cost up to 30,000 yuan. In the end, due to the scarce number of respondents who adopted this behaviour, we did not include this last option in the empirical model. These two strategies are quite different in nature from the previous behaviours, because they do not respond to pollution peaks and relate to long term perception of air pollution, rather than daily signals that can indicate an alert.

[15] Wage loss is computed multiplying the wage by the days at home, net of those covered by sick-leave: wage $\times$ (days lost - sick leave).

Information about air pollution. In our sample, internet use for the purpose of collecting information about air pollution is limited. The vast majority of people interviewed relied on government controlled sources of information, such as TV, radio or newspapers - see Table 8. Furthermore, 70 % of our sample considers the public information available on air pollution satisfactory and would not be interested in more information. Even after the worst pollution days in the previous year, only about a third started searching for more information about pollution peaks (Table 9). Therefore, the government can exert a direct influence through the media on the choices of a large fraction of the population when deciding what pollution signal to provide.

Starting from these stylized facts, we can analyse the determinants of averting behaviour and of the choice of the source of information.

4.3. Information

First of all, we characterize the agents that use government media as opposed to the internet. This can inform us about what groups are more affected by the government's signal. We set up a general empirical model of binary information choice, as follows:

$$I_h = \beta_0 + \mathbf{X_i}\beta + \mathbf{X_h}\gamma + \delta_3 W_h + \varepsilon_{it} \tag{16}$$

where I is the chosen mean of information, $\mathbf{X_i}$ is a vector of individual characteristics, $\mathbf{X_h}$ a vector of household characteristic, estimated as a probit. The individual controls are age, gender, education level and smoking status (to capture somehow risk attitude towards pollution and lung diseases). At the household level these are income and household size. Such a specification may pose some problems in terms of reverse causality, so we prefer a parsimonious model that uses only exogenous observables. Overall, this specification requires caution in making any causal statement, since there could be biases arising also from potential omitted variables.

First of all, we consider as left-hand-side variable the binary choice between using or not government controlled media to check air pollution information. The dependent variable equals to 1 if the respondent said that (s)he use TV, radio or newspapers as a source of information, and zero otherwise (so if he uses internet or no information about pollution at all, or self-perception). Then we distinguish between government media versus internet users, and government media versus self-perception users, to see what is driving the results. We also look at internet users versus everyone else.[16]

[16]For the determinants of internet, since the occurrence of a 1 is quite rare, we use a complementary log-log specification to account for the asymmetric nature of the dependent variable (Hilbe (1996)).

The results are displayed in Table 10. People who use government media are generally older, and this is driven by the internet users, who are typically young people. Those who prefer to use self-perception are less educated, but there is no significant difference in educational attainments between internet users and those who rely on government media. Interestingly, smokers declare to use the internet more than anything else. Larger households (families) rely a lot more on government sources than on the internet. Finally, income is strongly correlated with internet usage. This gives us an initial picture of those people who typically rely on the government's information.

4.4. Averting behaviour

In order to analyse self-protective behaviours in response to the government signal, we use a treatment-effect model, where the treatment is government information. The standard problem of missing observations is present in this case, as we cannot observe individual outcomes both under the treatment and non. We thus consider the average treatment effect (ATE), but given the non-random assignment of treatment we need to account for selection bias - as people choose what signal they want to listen to.

In order to identify our model, we consider only within-individual changes, namely how a person responded to different air pollution levels. Since many of the unobserved characteristics of a person remain constant under different pollution situations, this reduces the problem of omitted variables. Then we apply a two steps procedure, a bi-probit model - see Greene (2012), pp. 738-752 and Pindyck and Rubinfeld (1998), to see how the behaviour in response to pollution announcements differs between those people who are more or less likely to update their expectations and correct for the bias.

In order to do so, we include among the determinants of information sources in the first stage (Eq. 18) a dummy variable, S, which identifies with the value of 1 those people who consider the information they have sufficient to understand the quality of the air (and zero otherwise). This element captures the fraction of the population that is *less* likely to look for further information about pollution to correct for any possible information bias. It corresponds to the fraction of people, λ, defined in the theoretical model as those households who do not update expectations correcting for potential biases. In the household survey, this corresponds to those respondents who answered that information was enough to the question "*Do you think this [your current choice of] information is enough for you or would you like more of it? Specify the channel you would like to use more (TV, radio, newspaper, internet, other)*". Of course, those respondents who indicated they had enough information were not all users of government media as principal source of air quality news. In our case, we are interested in those people that for various cultural reasons consider the information they have as sufficient, and have chosen public media as the main source of news.

The specification of the two steps empirical model is the following

$$B_{ij} = \alpha_0 + \mathbf{X_i}\alpha_{h1} + \mathbf{X_h}\alpha_{h2} + \alpha_3 Z_h + \eta_{it} \tag{17}$$

$$Z_h = \beta_0 + \mathbf{X_i}\beta_{i1} + \mathbf{X_h}\beta_{h2} + \beta_3 S_h + \varepsilon_{it} \tag{18}$$

Behaviours, B, vary over individuals, i and over four possible activities, $j \in \{$*masks, transport, time outdoor, health checks*$\}$. With the exception of the last one, the dependent variable is measured in changes, taking the value of 1 if a person switches to more averting behaviour in extremely polluted days compared to normal days. Z is the choice of public information means, same as the previous specification: it takes the value of 1 when a person uses as principal source of information government-controlled public media (TV, radio, newspapers). Beyond the controls used previously, we add other controls: a dummy for workers, to distinguish individuals with more time flexibility from those with less; a control for car ownership in the transport specification, which may be particularly important as a sunk investment in averting; and a dummy for households with children, which could be possibly more careful about the health damages of pollution.

4.5. Results

Table 11 and 12 shows that, first of all, the first stage probit predicts quite well whether a person uses publicly controlled public media: those people who consider information sufficient are significantly more likely to choose government media, controlling for other factors. Then those people who use TV, radio and newspapers controlled by the Chinese Communist Party, and rely on it fully as identified by the first stage, are less likely to switch to more averting behaviours during peak pollution days, both in terms of time spent outdoors and for wearing masks. In the case of transport switch (third column), information does not play a significant role, but we can see that car ownership is a very strong determinant of this behaviour. It is harder for people to adopt safer means of transport if they do not own a car. In this case, it is not the information signal and how much people can update expectations on it that matters, but the investment in an asset as a car.

A different story applies to preventive health checks: those do not respond to the information signal negatively, but on the contrary correlate positively and significantly to it. This is not surprising, however: preventive checks are not just a response to peak pollution days, but rather an ex-ante self-protective behaviour. Moreover, public servants tent to have access to better health insurance in China and thus might do more preventive medical controls, even if they receive systematically biased air quality news.

Naturally this evidence is not conclusive about households trust in the government announcements, and possible people could have other mechanisms that they consider sufficient to understand pollution levels, even when receiving the news from government controlled media. However, overall these simple results fit well with the analysis of the previous section, which suggested that the government was reporting optimistically low values for pollution: for people who rely on such information and consider it sufficient for their choices, clearly they will not be alerted in high pollution days, and therefore they will be less likely to increase self-protective, even those ones that they could easily modify, such as spending less time outdoor and wearing masks.

5. Conclusion

A government with discretion over both pollution and information has incentives to control them jointly, in order to maximize the net profits from pollution in production and public support. Public information control enables the government to strike its own balance regarding the extent which workers will respond to the pollution from production. When a government has substantial control over information, and when it places a lower weight on health costs than production-sourced profits, then it will perceive a benefit from introducing a significant negative bias into information flows regarding pollution. This dampens the responsiveness of the working public to pollution, and shifts the costs of pollution so that they are experienced more as health damages than as production losses.

In this paper we examined these propositions in an empirical analysis of the announced and experienced air quality in Beijing, China. We analysed how the Beijing government provides public information about air pollution, and we found that the public signal is significantly downward biased as pollution increases. This bias is more pronounced around some critical thresholds, confirming the hypothesis that the government manipulates information in order to affect the perception of pollution. As a result, those urban dwellers who rely on government-controlled media adopt fewer measures to protect themselves during pollution peaks. Thus, the case of Beijing demonstrates that it can be optimal for a state with some control over both pollution and information to introduce bias within its information signals, but at the cost of inducing less risk-averting behaviour.

As the case of Beijing illustrates, the incentives for a government to provide a public good, such as the abatement of pollution, are lower when it can exercise control over the media. This has significant implications: in some developing countries, the problem with public goods' provision might not be the lack of capacity of the state, but rather the opposite, the excessive control of the government over certain aspects of the economy (such as information). In this system, this cycle of production-sourced pollution, distorted information, and reduced public responsiveness is a self-reinforcing reality, similar to a snake, or in this case a dragon, eating its own tail.

Bibliography

Alberini, A., Cropper, M., Fu, T.-T., Krupnick, A., Liu, J.-T., Shaw, D. and Harrington, W. (1997). Valuing health effects of air pollution in developing countries: The case of Taiwan, *Journal of Environmental Economics and Management* **34**(2): 107–126.

Andrews, S. Q. (2008). Inconsistencies in air quality metrics: 'Blue Sky' days and PM10 concentrations in Beijing, *Environmental Research Letters* **3**(3).

Angeletos, G.-M. and Pavan, A. (2007). Efficient use of information and social value of information, *Econometrica* **75**(4): 1103 –1142.

Asian Development Bank (2007). Country environmental analysis for the People's Republic of China, *ADB - Country Environmental Analysis Series* .

Aunan, K. and Pan, X. C. (2004). Exposure-response functions for health effects of ambient air pollution applicable for China: a meta-analysis, *Science of The Total Environment* **329**(1–3): 3 – 16.

Baeriswyl, R. and Cornand, C. (2010). The signaling role of policy actions, *Journal of Monetary Economics* **57**(6): 682 – 695.

Bayer, P., Keohane, N. and Timmins, C. (2009). Migration and hedonic valuation: The case of air quality, *Journal of Environmental Economics and Management* **58**(1): 1 – 14.

Besley, T. and Burgess, R. (2002). The political economy of government responsiveness: Theory and evidence from India, *Quarterly Journal of Economics* **117, No. 4**: 1415–1451.

Besley, T. and Ghatak, M. (2006). Public goods and economic development In: Banerjee, A., Benabou, R. and Mookherjee, D., (eds.) *Understanding Poverty*. Oxford University Press, Oxford.

Besley, T. and Prat, A. (2006). Handcuffs for the grabbing hand? Media capture and government accountability, *American Economic Review* **96**(3): 720–736.

Braga, A. L., Saldiva, P. H., Pereira, L. A., Menezes, J. J., Conceição, G. M., Lin, C. A., Zanobetti, A., Schwartz, J. and Dockery, D. W. (2001). Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil, *Pediatric Pulmonology* **31**(2): 106–113.

Brunetti, A. and Weder, B. (2003). A free press is bad news for corruption, *Journal of Public Economics* **87**(78): 1801 – 1824.

Chay, K. Y. and Greenstone, M. (2005). Does air quality matter? evidence from the housing market., *Journal of Political Economy* **113**(2): 376–424.

Chen, Y., Zhe, J. G., Naresh, K. and Guang, S. (2012). Gaming in air pollution data? Lessons from China, *The B.E. Journal of Economic Analysis and Policy* **13**(3): 1–43.

Clark, D. E., Herrin, W. E., Knapp, T. A. and White, N. E. (2003). Migration and implicit amenity markets: does incomplete compensation matter?, *Journal of Economic Geography* **3**(3): 289 – 307.

Congleton, R. D. (1992). Political institutions and pollution control, *Review of Economics and Statistics* **74**(3): pp. 412–421.

Cook, P. J. and Graham, D. A. (1977). The demand for insurance and protection: The case of irreplaceable commodities, *Quarterly Journal of Economics* **91**(1): pp. 143–156.

Deacon, R. (2009). Public good provision under dictatorship and democracy, *Public Choice* **139**(1-2): 241–262.

DellaVigna, S. and Kaplan, E. (2007). The Fox News effect: Media bias and voting, *Quarterly Journal of Economics* **122 (3)**(12169): 1187–1234.

Djankov, S., McLeish, C., Nenova, T. and Shleifer, A. (2003). Who owns the media?, *Journal of Law and Economics* **46**(2): 341–81.

Ehrlich, I. and Becker, G. S. (1972). Market insurance, self-insurance, and self-protection, *Journal of Political Economy* **80**(4): 623–48.

El-Fadel, M. and Massoud, M. (2000). Particulate matter in urban areas: health-based economic assessment, *Science of The Total Environment* **257**: 133–146.

Foster, A. and Kumar, N. (2011). Health effects of air quality regulations in Delhi, India, *Atmospheric Environment* **45**: 1675–1683.

Gavazza, A. and Lizzeri, A. (2009). Transparency and economic policy, *Review of Economic Studies* **76**(3): 1023–1048.

Ghanem, D. and Zhang, J. (2014). Effortless perfection: Do Chinese cities manipulate air pollution data?, *Journal of Environmental Economics and Management* **68**(2): 203 – 225.

Greene, W. H. (2012). *Econometric Analysis*, Prentice Hall, 7th ed. Upper Saddle River, NJ.

Hilbe, J. M. (1996). Maximum-likelihood complementary log-log regression., *Stata Technical Bulletin Reprints, Stata Press* **6**: 129–131.

Islam, R. (2006). Does more transparency go along with better governance?, *Economics and Politics* **18**(2): 121–167.

Jalan, J. and Somanathan, E. (2008). The importance of being informed: Experimental evidence on demand for environmental quality, *Journal of Development Economics* **87**(1): 14 – 28.

Kennedy, P. W., Laplante, B. and Maxwell, J. (1994). Pollution policy: the role for publicly provided information, *Journal of Environmental Economics and Management* **26**(1): 31 – 43.

Madajewicz, M., Pfaff, A., van Geen, A., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R. and Ahsan, H. (2007). Can information alone change behavior? Response to arsenic contamination of groundwater in Bangladesh, *Journal of Development Economics* **84**(2): 731 – 754.

Matus, K., Nam, K.-M., Selin, N. E., Lamsal, L. N., Reilly, J. M. and Paltsev, S. (2012). Health damages from air pollution in China, *Global Environmental Change* **22**(1): 55 – 66.

Mèon, P. G. and Minne, G. (2014). Mark my words: Information and the fear of declaring an exchange rate regime, *Journal of Development Economics* **107**(0): 244 – 261.

Michalski, T. and Stoltz, G. (2013). Do countries falsify economic data strategically? some evidence that they might, *The Review of Economics and Statistics* **95**(2): 591–616.

Milhaupt, C. J. and Zheng, W. (2014). Beyond ownership: State capitalism and the Chinese firm, *Georgetown Law Journal* **103**: (forthcoming).

Moretti, E. and Neidell, M. (2011). Pollution, health, and avoidance behavior: Evidence from the ports of Los Angeles, *Journal of Human Resources* **46**(1): 154–175.

Morris, S. and Shin, H. S. (2002). Social value of public information, *American Economic Review* **92**(5): 1521–1534.

Moscarini, G. (2007). Competence implies credibility, *American Economic Review* **97**(1): 37–63.

Persico, N. and Lizzeri, A. (2001). The provision of public goods under alternative electoral incentives, *American Economic Review* **91**(1): 225–239.

Pindyck, R. S. and Rubinfeld, D. L. (1998). *Econometric Models and Economic Forecasts*, McGraw-Hill,4th ed. New York.

Pittel, K. and Rubbelke, D. T. (2008). Climate policy and ancillary benefits: A survey and integration into the modelling of international negotiations on climate change, *Ecological Economics* **68**(1-2): 210–220.

Qiu, H., sun Yu, I. T., Tian, L., Wang, X., Tse, L. A., Tam, W. and Wong, T. W. (2012). Effects of coarse particulate matter on emergency hospital admissions for respiratory diseases: A time-series analysis in Hong Kong, *Environmental Health Perspectives* **120(4)**: 572 – 576.

Ravetti, C., Swanson, T., Jin, Y. P., Quan, M. and Shiquiu, Z. (2014). A household survey of the cost of illness due to air pollution in Beijing, China., *CIES Research Paper* **No. 28**.

Ren, C. and Tong, S. (2008). Health effects of ambient air pollution - recent research development and contemporary methodological challenges, *Environmental Health* **7**(1): 56.

Sakulniyomporn, S., Kubaha, K. and Chullabodhi, C. (2011). External costs of fossil electricity generation: Health-based assessment in Thailand, *Renewable and Sustainable Energy Reviews* **15**(8): 3470–3479.

Simmons, K. M. and Kruse, J. B. (2000). Market value of mitigation and perceived risk: Empirical results, *Journal of Economics* **26**(1): 41–51.

Snyder, J. M. and Strömberg, D. (2010). Press coverage and political accountability, *Journal of Political Economy* **118**(2): 355–408.

Somanathan, E. (2010). Effects of information on environmental quality in developing countries, *Review of Environmental Economics and Policy* .

Stiglitz, J. E. (2000). The contributions of the economics of information to twentieth century economics, *Quarterly Journal of Economics* **115**(4): 1441–1478.

Stoerk, T. (2015). Statistical corruption in Beijing's air quality data has likely ended in 2012, *Grantham Research Institute on Climate Change and the Environment* **Working Paper No. 194**.

Talberth, J., Berrens, R. P., Mckee, M. and Jones, M. (2006). Averting and insurance decisions in the wildland-urban interface: Implications of survey and experimental data for wildfire risk reduction policy, *Contemporary Economic Policy* **24**(2): 203–223.

US - Environmental Protection Agency (2006). Guideline for reporting of daily air quality - Air Quality Index (AQI), *EPA-454/B-06-001* **May 2006**.

Wang, W., Primbs, T., Tao, S. and Simonich, S. L. M. (2009). Atmospheric particulate matter pollution during the 2008 Beijing Olympics, *Environmental Science & Technology* **43**(14): 5314–5320.

Whitehead, J. (2005). Environmental risk and averting behavior: Predictive validity of jointly estimated revealed and stated behavior data, *Environmental & Resource Economics* **32**(3): 301–316.

WHO (2005). *Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Global update 2005. Summary of risk assessment.*, World Health Organization.

Williams, A. (2009). On the release of information by governments: Causes and consequences, *Journal of Development Economics* **89**(1): 124 – 138.

Zhang, P., Dong, G., Sun, B., Zhang, L., Chen, X., Ma, N., Yu, F., Guo, H., Huang, H., Lee, Y. L., Tang, N. and Chen, J. (2012). Long-term exposure to ambient air pollution and mortality due to cardiovascular disease and cerebrovascular disease in Shenyang, China, *Respiration* **84 (5)**(6): 360–368.

Zheng, S. and Kahn, M. E. (2013). Understanding china's urban pollution dynamics, *Journal of Economic Literature* **51**(3): 731–72.

Appendix

Optimal bias and pollution. The problem of the government in eq. (1) can be written as

$$\underset{(\theta,\beta)}{max}\ \ V = \pi(\theta, n(w) - kE(\theta)) - c[d\theta - aE(\theta)]\ [n(w) - kE(\theta)] \tag{19}$$

The government considers how people form their expectations about pollution and optimally responds to that. Following eq. (7), on average expectations about pollution in the long run, as $E(\beta) \to \beta$, are

$$E(\theta) = \frac{A-(1-\lambda)\beta}{z} = \frac{\theta+\lambda\beta}{z} \quad \text{where} \quad z \equiv W^2\left(\frac{1}{\sigma_\theta^2} + \frac{1}{W^2}\right)$$

Plugging it in the government maximization above, we find two first order conditions with respect to θ and β, which can then be solved simultaneously.

$$\frac{\partial V}{\partial \theta}:\ \theta = \frac{1}{2}\left(\frac{n(w)z}{k} + \frac{z^2}{ck(a-dz)}\pi_\theta - \frac{z}{(a-dz)c}\pi_N - \frac{2a-dz}{a-dz}\lambda\beta\right)$$

$$\frac{\partial V}{\partial \beta}:\ \beta = \frac{1}{2ca\lambda k}\left(kw + canw - \frac{\partial y}{\partial N}k - cdk\theta\right)$$

Plugging the first into the second we solve for β^*, a function of λ and all the various elements presented in the model.

$$\beta^* = \frac{1}{\lambda}\ \left(\frac{dz-a}{kd}\ n(w)\ -\ \frac{1}{cd}\ \pi_N\ -\ \frac{dz-2a}{ckd^2}\ \pi_\theta\right) \tag{20}$$

This is the same result as in Section 2.

Plugging this result in the first order condition for θ, we get the result in the paper for θ^*.

$$\theta^* = \frac{a}{kd}\ n(w) + \frac{1}{cd}\ \pi_N - \frac{2a}{ckd^2}\ \pi_\theta \tag{21}$$

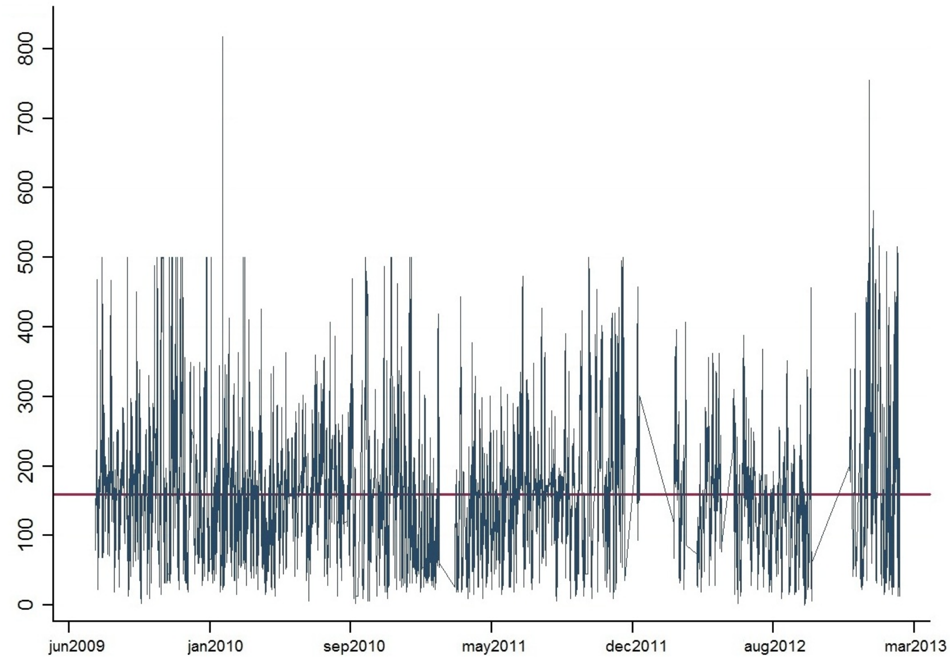

FIGURE 3: US HOURLY AIR POLLUTION INDEX

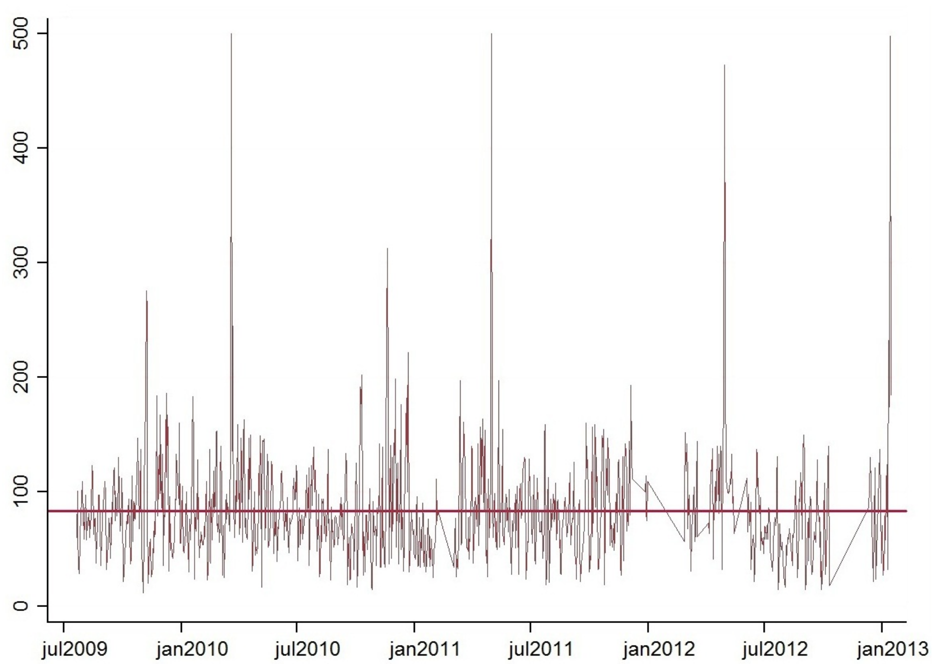

FIGURE 4: CHINESE DAILY AIR POLLUTION INDEX

TABLE 1: AIR POLLUTION INDEXES

Index and Definition			Health Implications	PM10 ($\mu g/m3$)		NOx ($\mu g/m3$)	
AQI US	API China			US	China	US	China
0-50 Good	0-50 Excellent		Air quality is considered satisfactory and air pollution poses little or no risk.	0-50	0-50	0-0.03	0-50
51-100 Moderate	51-100 Good		Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.	50-150	50-150	0.03 - 0.14	50-150
101-150 Unhealthy for sensitive groups	100-200	Slightly polluted	Members of sensitive groups may experience health effects. The general public is not likely to be affected.	150-250	150-350	0.14 - 0.22	150-800
151-200 Unhealthy		Light polluted	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.	250-350		0.22 - 0.30	
201-300 Very Unhealthy	200-300	Moderately polluted Moderate-heavy polluted	Health warnings of emergency conditions. The entire population is more likely to be affected.	350 - 420	350-420	0.30 - 0.60	800-1600
300+ Hazardous	300-400	Heavy polluted	Health alert: everyone may experience more serious health effects.	420-600	420-500	0.60 - 1.0	1600-2100
	400-500				500-600		2100-2620
	500				600		2620

Source: own elaboration from US Environmental Protection Agency and China's Ministry of Environmental Protection.

TABLE 2: SUMMARY STATISTICS

Variable	Obs	Mean	Std. Dev.	Min	Max
Chinese Index	1083	82.9	45.9	12	500
US index (daily average)	1083	158.6	78.2	9.1	528.1
US index (daily minimum)	1083	103.3	69.9	0	500
US index (daily maximum)	1083	230.5	109.9	13	817
Log Chinese Index	1083	4.3	.5	2.5	6.2
Log US index (daily average)	1083	4.9	.6	2.2	6.3
Log US index (daily minimum)	1082	4.3	.8	1.1	6.2
Log US index (daily maximum)	1083	5.3	.5	2.6	6.7
DEPENDENT VARIABLE					
Log Chinese /US index (avg)	1083	-.6	.3	-1.8	1.7
Log Chinese /US index (min)	1082	-.1	.7	-1.3	2.9
Log Chinese /US index (max)	1083	-1.1	.4	-2.6	1.7
THRESHOLD DUMMIES					
Threshold 100 points	1083	.78		0	1
Threshold 200 points	1083	.27		0	1
Threshold 300 points	1083	.09		0	1

TABLE 3: CAUSES OF THE DISCREPANCY BETWEEN THE CHINESE AND US INDEX

	Dependent variable: Gap China-US signal (daily average)			
	(1)	(2)	(3)	(4)
US Index	−0.79***	−0.79***	−0.79***	−0.79***
	(0.02)	(0.02)	(0.02)	(0.02)
Threshold 100	−1.58***	−1.59***	−1.53***	−1.60***
	(0.30)	(0.30)	(0.31)	(0.30)
Threshold 200	0.10	0.03	0.15	0.07
	(0.89)	(0.90)	(0.91)	(0.90)
Threshold 300	−2.14**	−2.02*	−2.35**	−2.07*
	(1.05)	(1.06)	(1.06)	(1.06)
T100*US Index	0.33***	0.33***	0.32***	0.34***
	(0.06)	(0.06)	(0.06)	(0.06)
T200*US Index	0.00	0.02	−0.01	0.01
	(0.16)	(0.17)	(0.17)	(0.17)
T300*US Index	0.38**	0.36*	0.41**	0.36*
	(0.19)	(0.19)	(0.19)	(0.19)
Constant	3.05***	3.04***	3.04***	3.06***
	(0.14)	(0.14)	(0.14)	(0.14)
L.ar	0.34***	0.32***		
	(0.03)	(0.03)		
L2.ar		0.05		
		(0.03)		
L.ma			0.30***	0.31***
			(0.03)	(0.03)
L2.ma				0.14***
				(0.03)
Constant	0.21***	0.21***	0.21***	0.21***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	876	876	876	876
AIC	−197.63	−197.29	−183.31	−193.82
BIC	−73.47	−68.35	−59.15	−64.89

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The lower panel shows the autoregressive moving average lags (ARMA) components.

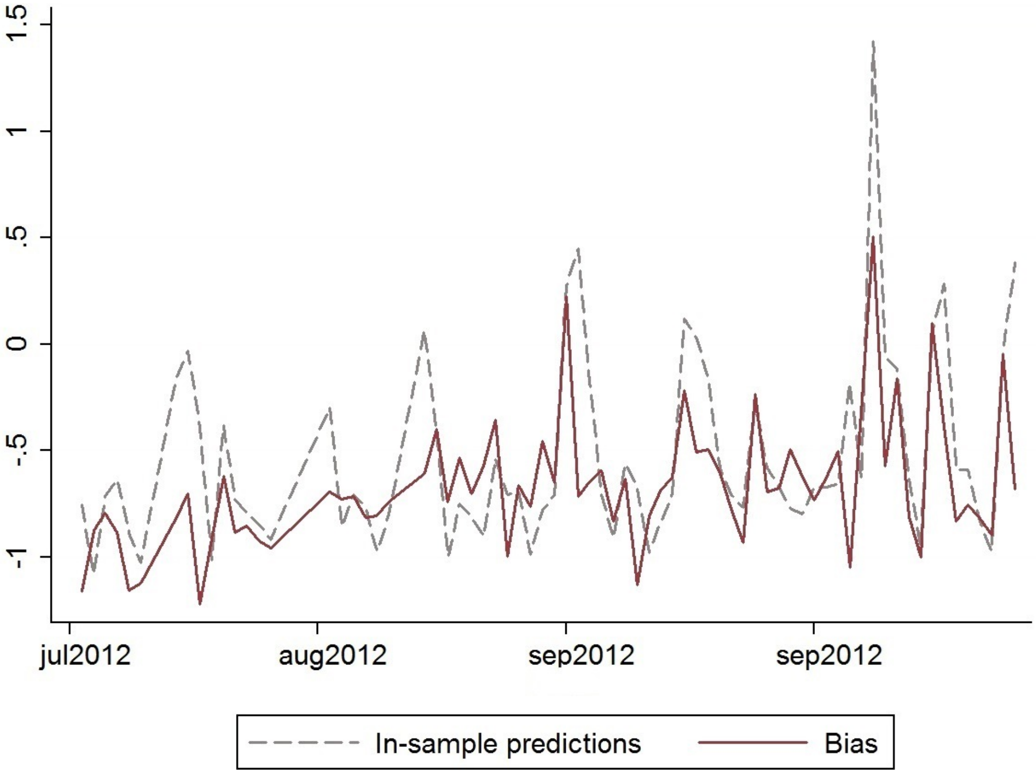

FIGURE 5: MODEL PERFORMANCE IN PREDICTING THE BIAS

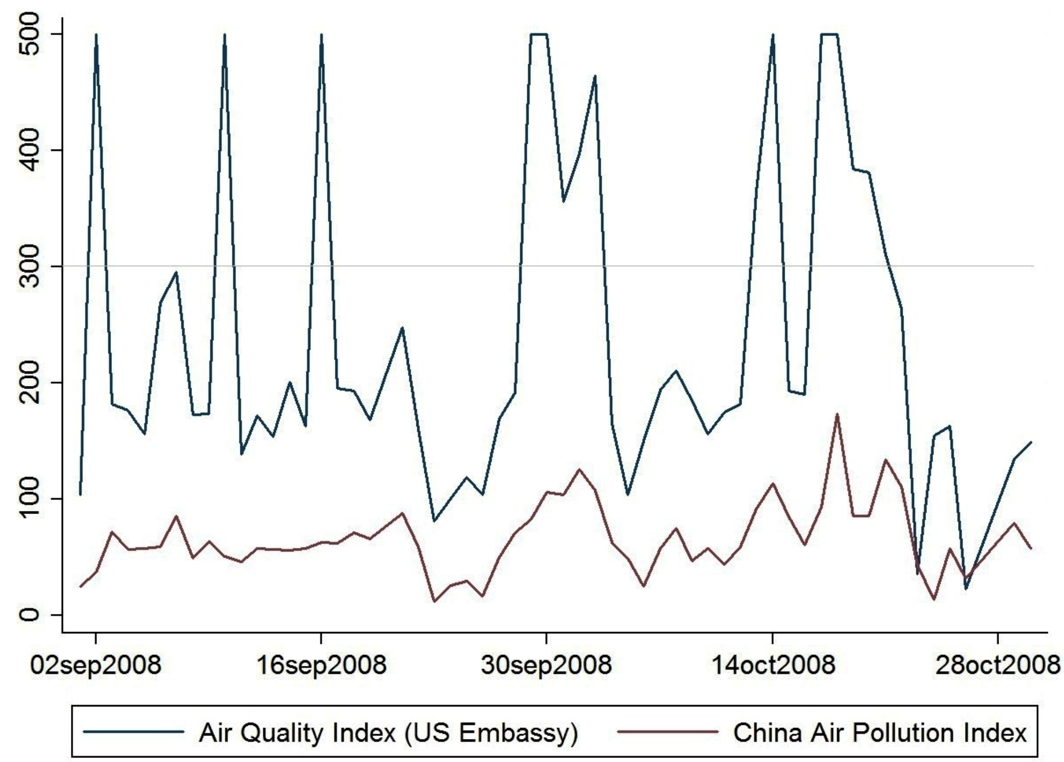

FIGURE 6: INDEX MISMATCH WITH DAILY MAXIMA

TABLE 4: CHINESE/USA AIR POLLUTION INDEXES (MINIMUM OF US HOURLY OBSERVATIONS)

	(1)	(2)	(3)	(4)
Minimum AQI	−0.92***	−0.92***	−0.92***	−0.92***
	(0.01)	(0.01)	(0.01)	(0.01)
AQI threshold(100)	−1.86***	−1.86***	−1.81***	−1.86***
	(0.29)	(0.30)	(0.30)	(0.30)
AQI threshold(200)	0.41	0.35	0.47	0.37
	(0.87)	(0.87)	(0.88)	(0.88)
AQI threshold(300)	−2.56**	−2.48**	−2.78***	−2.52**
	(1.04)	(1.05)	(1.05)	(1.05)
T100*avg. AQI	0.40***	0.40***	0.39***	0.40***
	(0.06)	(0.06)	(0.06)	(0.06)
T200*avg. AQI	−0.05	−0.04	−0.06	−0.05
	(0.16)	(0.16)	(0.16)	(0.16)
T300*avg. AQI	0.45**	0.44**	0.49***	0.44**
	(0.18)	(0.19)	(0.19)	(0.19)
Constant	3.69***	3.69***	3.68***	3.70***
	(0.11)	(0.11)	(0.10)	(0.11)
L.ar	0.34***	0.32***		
	(0.03)	(0.03)		
L2.ar		0.04		
		(0.03)		
L.ma			0.30***	0.32***
			(0.03)	(0.03)
L2.ma				0.12***
				(0.03)
Constant	0.21***	0.21***	0.21***	0.21***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	876	876	876	876
AIC	−159.64	−158.62	−147.24	−155.81
BIC	−35.48	−29.68	−23.08	−26.87

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The lower panel shows the autoregressive moving average lags (ARMA) components.

TABLE 5: CHINESE/USA AIR POLLUTION INDEXES (MAXIMUM OF US HOURLY OBSERVATIONS)

	(1)	(2)	(3)	(4)
Maximum AQI	−0.82***	−0.82***	−0.82***	−0.83***
	(0.02)	(0.02)	(0.02)	(0.02)
AQI threshold(100)	−1.65***	−1.66***	−1.61***	−1.67***
	(0.30)	(0.31)	(0.31)	(0.31)
AQI threshold(200)	0.17	0.14	0.25	0.18
	(0.89)	(0.90)	(0.91)	(0.90)
AQI threshold(300)	−1.98*	−1.91*	−2.25**	−1.98*
	(1.04)	(1.04)	(1.05)	(1.04)
T100*avg. AQI	0.35***	0.35***	0.34***	0.36***
	(0.06)	(0.06)	(0.06)	(0.06)
T200*avg. AQI	−0.01	−0.01	−0.03	−0.02
	(0.16)	(0.17)	(0.17)	(0.17)
T300*avg. AQI	0.34*	0.33*	0.39**	0.34*
	(0.18)	(0.19)	(0.19)	(0.18)
Constant	3.08***	3.07***	3.08***	3.08***
	(0.15)	(0.16)	(0.14)	(0.15)
L.ar	0.36***	0.35***		
	(0.03)	(0.03)		
L2.ar		0.03		
		(0.03)		
L.ma			0.33***	0.34***
			(0.03)	(0.03)
L2.ma				0.14***
				(0.03)
Constant	0.21***	0.21***	0.21***	0.21***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	876	876	876	876
AIC	−157.92	−156.43	−142.55	−154.78
BIC	−33.76	−27.50	−18.39	−25.85

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The lower panel shows the autoregressive moving average lags (ARMA) components.

A2 - Household Survey

TABLE 6: PRIVATE COST OF ILLNESS

	Direct costs	Days of work lost	Paid sick leave	Days of inactivity	Indirect costs	Total
Airborne diseases	2514 yuan	1.4	0.5	9	812 yuan	3326 yuan
All illnesses	5184 yuan	18	13	53	305 yuan	5489 yuan

TABLE 7: CHARACTERISTICS OF AVERTING BEHAVIOURS

	Frequency	Observations
Everyday's life		
Cancel outdoor activities	58%	1621
Change means of transportation	6%	1602
Wear a mask	11%	1618
Peak pollution times		
Cancel outdoor activities	77%	1240
Change means of transportation	12%	1231
Wear a mask	18%	1245
Change of behaviour		
Reduce time outdoor	23%	1239
Transport change	5%	1224
Mask	9%	1238
Other behaviours		
Preventive health checks	73%	1626
Air purifier	23%	1639

TABLE 8: DIFFERENT SOURCES OF INFORMATION ABOUT AIR POLLUTION

	Frequency
Government sources (TV, radio, newspapers)	77%
Internet (PC or mobile device)	6%
Self-perception, other people	17%
Doesn't care	0.1%

TABLE 9: WHAT DID YOU DO AFTER THE PEAK POLLUTION DAYS LAST YEAR?

Nothing	38%
I started worrying more about air pollution	25%
I look for more information	9%
I worry more about air pollution and look for more information about it	27%
Other	1%

TABLE 10: SOURCES OF INFORMATION

	Government media	Govt vs. Internet	Govt vs. Self	Internet
Age	0.01***	0.04***	0.00	–0.03***
	(0.00)	(0.01)	(0.00)	(0.01)
Male	0.06	0.13	0.03	–0.15
	(0.10)	(0.17)	(0.11)	(0.15)
Education	0.15**	–0.08	0.22***	0.13
	(0.06)	(0.12)	(0.07)	(0.11)
Smoker	–0.48***	–0.69***	–0.41*	0.56**
	(0.18)	(0.27)	(0.21)	(0.25)
Migrant	–0.46	0.16	–0.61	–0.30
	(0.35)	(0.89)	(0.37)	(0.80)
Household size	0.25**	0.69**	0.12	–0.67**
	(0.11)	(0.32)	(0.12)	(0.29)
Household Income	–2.13	–5.66***	–0.10	5.45***
	(1.45)	(1.93)	(1.90)	(1.63)
Constant	–0.76	–0.51	–0.04	–0.13
	(0.50)	(0.96)	(0.57)	(0.84)
Observations	1490	1260	1408	1490

Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 11: SELF-PROTECTIVE BEHAVIOURS AND INFORMATION (BI-PROBIT)

	Outdoor Δ	Mask Δ	TransportΔ	Health checks
Smoker	−0.21*	−0.46**	0.04	−0.10
	(0.13)	(0.20)	(0.15)	(0.11)
Worker	−0.06	0.35*	0.21	−0.06
	(0.14)	(0.20)	(0.24)	(0.12)
Children	0.17	0.29	−0.18	0.28*
	(0.15)	(0.18)	(0.24)	(0.15)
Household Income	−0.05	−0.33	0.06	3.09***
	(1.02)	(1.08)	(1.35)	(0.73)
Public Media	−1.37***	−2.18***	−0.45	1.73***
	(0.48)	(0.56)	(0.50)	(0.29)
Car	–	–	0.57***	–
			(0.21)	
	Government media			
Sufficient info	0.39**	0.37**	0.38**	0.29**
	(0.16)	(0.17)	(0.17)	(0.13)
Smoker	−0.34**	−0.33**	−0.29**	−0.28**
	(0.13)	(0.13)	(0.13)	(0.12)
Worker	0.06	0.06	0.09	0.08
	(0.16)	(0.16)	(0.17)	(0.14)
Children	−0.22	−0.17	−0.17	−0.02
	(0.20)	(0.20)	(0.21)	(0.18)
Household Income	−0.15	0.16	−0.42	−1.15
	(1.10)	(1.08)	(1.03)	(0.77)
Observations	1147	1146	1133	1428

Clustered standard errors (household) in brackets. * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$). Respondent, Age, Male, Education controls and Constants not reported.

TABLE 12: SELF PROTECTIVE BEHAVIOURS AND INFORMATION (IV-PROBIT)

	Outdoor (Δ)	Mask (Δ)	Transport(Δ)	Health checks	Air purifier
Government media	−2.25***	−2.74***	−2.01*	−1.42	1.79
	(0.86)	(0.44)	(1.06)	(1.23)	(1.29)
Smoker	−0.21*	−0.40**	0.04	−0.20*	0.08
	(0.11)	(0.17)	(0.15)	(0.12)	(0.12)
Worker	−0.01	0.31*	0.22	−0.06	0.15
	(0.13)	(0.19)	(0.21)	(0.14)	(0.16)
Children	0.03	0.11	−0.37*	0.22	0.57**
	(0.16)	(0.19)	(0.21)	(0.17)	(0.27)
Household Income	0.04	−0.21	−0.77	2.67**	1.31
	(1.00)	(0.82)	(1.21)	(1.12)	(0.82)
Car			0.63***		
			(0.19)		
	Government media				
Smoker	−0.06**	−0.05*	−0.06**	−0.05*	−0.05*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Worker	0.02	0.02	0.03	0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Children	−0.05	−0.04	−0.08*	−0.03	−0.03
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
Household Income	−0.16	−0.16	−0.33	−0.08	−0.08
	(0.26)	(0.27)	(0.29)	(0.23)	(0.23)
Sufficient info	0.08**	0.08**	0.09**	0.08**	0.08**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Car			0.10***		
			(0.04)		
athrho	0.82	1.22**	0.76	0.47	−0.76
	(0.54)	(0.52)	(0.59)	(0.50)	(0.64)
lnsigma	−1.12***	−1.12***	−1.13***	−1.11***	−1.11***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Street Controls	Yes	Yes	Yes	Yes	Yes
Observations	1103	1103	1093	1353	1356
Wald Test p value	0.13	0.02	0.20	0.34	0.24

Clustered standard errors in parentheses (household). $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$